

**SEGMENTACIÓN DE AFILIADOS A UN FONDO DE PENSIONES  
OBLIGATORIAS APLICANDO LA TÉCNICA DE CLASIFICACIÓN POR  
CONGLOMERADOS JERÁRQUICOS CON BASE EN ALGUNAS VARIABLES  
CUALITATIVAS**

**RICARDO GIL RUBIO  
CARLOS IVÁN UREÑA**

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES  
DEPARTAMENTO DE CIENCIAS BÁSICAS  
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA  
BOGOTÁ, D.C.**

**2015**

**SEGMENTACIÓN DE AFILIADOS A UN FONDO DE PENSIONES  
OBLIGATORIAS APLICANDO LA TÉCNICA DE CLASIFICACIÓN POR  
CONGLOMERADOS JERÁRQUICOS CON BASE EN ALGUNAS VARIABLES  
CUALITATIVAS**

**RICARDO GIL RUBIO  
CARLOS IVÁN UREÑA**

**TRABAJO DE GRADO PARA OPTAR EL TÍTULO DE  
ESPECIALISTA EN ESTADÍSTICA APLICADA**

**ANA MARÍA GÓMEZ LAMUS  
M. SC. ESTADÍSTICA**

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES  
DEPARTAMENTO DE CIENCIAS BÁSICAS  
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA  
BOGOTÁ, D.C.**

**2015**

Nota de aceptación

---

---

---

---

---

---

---

Presidente del Jurado

---

Jurado

---

Jurado

Bogotá D.C. 30 de Julio de 2015

### **Dedicado A**

Este trabajo de grado no habría sido posible sin la ayuda de mis padres, que han estado prestos a colaborarme en todo momento, que han sido pacientes porque no les he dedicado el tiempo necesario y que han entendido la priorización a este proyecto para hacer posible su culminación. A Dios que ha permitido que esté bien de salud, con todos los sentidos puestos en este maravilloso trabajo y que en aflicciones o desmotivación a estado irradiando buena energía para encontrar el camino correcto.

**Ricardo Gil Rubio.**

A Dios y la Virgen santísima por darme la vida y salud. A toda mi familia, en especial a mis padres Iván E. Ureña y Luz Marina Molina, mi hermana María del Pilar Ureña Molina, quienes con su apoyo incondicional hicieron realidad este proyecto profesional tan importante en mi vida. Este logro es dedicado a ustedes.

**Carlos I. Ureña Molina**

## **AGRADECIMIENTOS**

El presente trabajo de grado fue realizado bajo la supervisión de la M. Sc. Ana María Gómez Lamus, a quien queremos agradecer profundamente por el apoyo brindado, por su paciencia, tiempo y dedicación que tuvo para que resultará de forma satisfactoria el presente trabajo de grado. A nuestros padres por su apoyo en cada momento de nuestras vidas para culminar los proyectos que nos hemos trazado. A mis maestros que aportaron todo su conocimiento, tiempo, dedicación y su pasión por la estadística y la actividad docente.

## CONTENIDO

	PÁG.
1. INTRODUCCIÓN.....	14
2. PROBLEMA.....	15
2.1 PLANTEAMIENTO DEL PROBLEMA.....	15
2.2 FORMULACIÓN DEL PROBLEMA.....	15
3. OBJETIVO.....	16
3.1 OBJETIVO GENERAL.....	16
3.2 OBJETIVO ESPECIFICO.....	16
4. JUSTIFICACIÓN.....	17
5. MARCO TEÓRICO.....	18
5.1 Segmentación de Mercado.....	18
5.2 Análisis de Correspondencias (ACM).....	19
5.3 Análisis de Conglomerados Jerárquicos.....	21
5.4 Tipos de Métodos de Clasificación por Conglomerados Jerárquicos....	22
5.5 Identificación de Variables Objeto de Estudio.....	26
6. MARCO LEGAL.....	27
7. MARCO METODOLÓGICO.....	28
7.1 Datos de Estudio.....	28
7.2 Tipo de Investigación – Análisis Descriptivo.....	28
7.3 Variables de Estudio.....	29
7.4 Recolección de Información.....	30

7.5	Diseño Estadístico.....	32
8.	RESULTADOS.....	33
8.1	Análisis Descriptivo.....	33
8.1.1	Descripción Tablas de Contingencia.....	33
8.1.2	Análisis de Correspondencia Múltiple (ACM).....	34
8.2	Aplicación y Selección de la Técnica de Clasificación Jerárquica.....	38
8.3	Dendrograma resultante de la Técnica de Clasificación Jerárquica con el Método de Ward.....	39
8.4	Resultados de la Técnica de Clasificación Jerárquica Método de Ward.....	40
9.	CONCLUSIONES.....	46
10.	DISCUSIÓN.....	48
11.	LISTA DE REFERENCIAS.....	49

## LISTA DE TABLAS

Tabla 1. Mercado Orientado al Consumidor y a la Industria.....	19
Tabla 2. Identificación de Variables Objeto de Estudio .....	26
Tabla 3. Normas que Rigen la Seguridad Social en Colombia .....	27
Tabla 4. Variables Cualitativas.....	29
Tabla 5. Estadísticos Descriptivos IBC entre 4.284.801 y 8.569.600 – IBC entre 2.142.401 y 4.284.800 .....	33
Tabla 6. Estadísticos Descriptivos IBC entre 1.071.201 y 2.142.400 – IBC entre 535.600 y 1.071.200.....	34
Tabla 7. Resumen del Modelo.....	35
Tabla 8. Medidas de Discriminación.....	36
Tabla 9. Descripción de Variables Según Diagrama Conjunto de Puntos de Categoría .....	38
Tabla 10. Clasificación Jerárquica Grupo con 3 Conglomerados .....	38
Tabla 11. Clasificación Jerárquica Grupo con 4 Conglomerados .....	39
Tabla 12. Clasificación Jerárquica Grupo con 5 Conglomerados .....	39
Tabla 13. Conglomerados Jerárquicos (3) para el Método de Ward .....	40
Tabla 14. Análisis Conglomerados Jerárquicos (conglomerado 1) .....	41
Tabla 15. Análisis Conglomerados Jerárquicos (conglomerado 2) .....	42
Tabla 16. Análisis Conglomerados Jerárquicos (conglomerado 3) .....	43
Tabla 17. Resumen Análisis de Conglomerados Jerárquicos 1,2 y 3. ....	44
Tabla 18. Estrategias Comerciales .....	48



## LISTA DE FIGURAS

<b>Figura 1. Vecino más Lejano.....</b>	<b>22</b>
<b>Figura 2. Vecino más Cercano.....</b>	<b>23</b>
<b>Figura 3. Vinculación Intergrupos .....</b>	<b>23</b>
<b>Figura 4. Vinculación Intragrupos .....</b>	<b>23</b>
<b>Figura 5. Método Centroide .....</b>	<b>24</b>
<b>Figura 6. Método de Ward .....</b>	<b>25</b>
<b>Figura 7. Metodología KDD .....</b>	<b>28</b>
<b>Figura 8. Procedimiento de Recolección de Información.....</b>	<b>30</b>
<b>Figura 9. Proceso Estadístico.....</b>	<b>32</b>
<b>Figura 10. Escala de consistencia interna – Alfa de Cronbach .....</b>	<b>35</b>
<b>Figura 11. Medidas de Discriminación .....</b>	<b>36</b>
<b>Figura 12. Diagrama Conjunto de Puntos de Categorías.....</b>	<b>37</b>
<b>Figura 13. Dendograma Método de Ward grupo con 3 conglomerados .....</b>	<b>40</b>

## GLOSARIO

**AFP:** administradoras de fondos de pensiones, son sociedades financieras constituidas de acuerdo a las leyes del país, con el objeto exclusivo de administrar las cuentas personales de los afiliados e invertir adecuadamente en los fondos.

**AS400:** sistema fuente de información que contiene los módulos de Pensiones Obligatorias, Cesantías y Pensiones Voluntarias.

**Dendograma:** es un tipo de representación gráfica o diagrama de datos en forma de árbol que organiza los datos en subcategorías que se van dividiendo en otros hasta llegar al nivel detallado y deseado (asemejándose a las ramas de un árbol que se van dividiendo en otras sucesivamente).

**Error Cuadrático Medio (ECM):** es una forma de evaluar la diferencia entre un estimador y el valor real de la cantidad que se quiere calcular. El ECM mide el promedio del cuadrado del error, siendo el error el valor en la que el estimador difiere de la cantidad hacer estimada.

**Ingreso Base de Cotización (IBC):** es el monto del salario sobre el cual se aplica el porcentaje de cotización a pensión. Es importante tener en cuenta que la Ley dispuso como tope máximo de IBC para todos los trabajadores de 25 salarios mínimos mensuales legales vigentes. Para trabajadores bajo el régimen de salario integral (más de 10 SMMLV) el IBC corresponde al 70% del salario siempre y cuando ese valor no exceda el tope Máximo de 25 salarios mínimos.

**Knowledge Discovery in Datamining (KDD):** que se refiere al proceso no-trivial de descubrir conocimiento e información potencialmente útil dentro de los datos contenidos en algún repositorio de información.

**Mercado:** gente o empresas con necesidades o deseos y la capacidad o voluntad de comprar.

**Pensiones Obligatorias (PO):** es el ahorro obligatorio que deben hacer los trabajadores Colombianos dependientes o independientes a lo largo de su vida laboral para asegurar un ingreso mensual en la vejez. Este ahorro también actúa

como un seguro ante la invalidez o la muerte prematura.

**PQR:** peticiones, quejas y reclamos que radica un afiliado o empleador a un fondo de pensiones para que le den solución a su solicitud.

**Segmentar:** el término se refiere al proceso de definición y división de un amplio mercado en grupos claramente identificables y homogéneos de consumidores con necesidades, deseos y características similares.

## RESUMEN

El siguiente trabajo de grado se realizó con el propósito de segmentar los afiliados de un fondo de pensiones obligatorias en Colombia teniendo en cuenta las siguientes variables: edad, género, IBC, tipo de empresa, tipo de afiliado (PO) y regional.

Lo anterior se realizó con la implementación de una técnica de estadística computacional por conglomerados jerárquicos, siguiendo la metodología de investigación Knowledge Discovery in Datamining (KDD) con el fin de conocer información potencialmente útil dentro de los datos.

El procesamiento de la información se realizó con el software estadístico SPSS (versión 21), se aplicó la técnica de clasificación por conglomerados jerárquicos en un conjunto de entrenamiento de 500 casos y se obtuvo el resultado de la clasificación para grupos de 3, 4 y 5 conglomerados utilizando los métodos intergrupos, intragrupos, vecino más cercano, vecino más lejano, agrupación de medianas, agrupación de centroides y método de Ward.

Para la identificación de la mejor segmentación se tuvo en cuenta en las técnicas de clasificación, la comparación del error cuadrático medio, la mejor distribución de los datos de acuerdo al criterio de un experto y las buenas características del

método seleccionado según la teoría, estas características se presentaron para el grupo con 3 conglomerados en el método de Ward.

Posteriormente se realizó una caracterización de los casos agrupados en cada uno de los 3 conglomerados, encontrando que las variables que más se asocian a la dimensión 1 son tipo de afiliado PO, tipo de empresa, IBC, mientras que las variables que se asocian a la dimensión 2, edad, género y regional.

### **ABSTRACT**

The following thesis was conducted with the purpose of segmenting members of a mandatory pension fund in Colombia taking into account the following variables: age, gender, IBC, type of business, type of affiliated member (PO) and the region. (District)

It was done with the implementation of a computational statistical technique classified into hierarchical clusters. Inspired by the methodology of research in “Knowledge Discovery Data mining (KDD)” in order to meet potentially useful information within the data.

The processing of the information was performed using the statistical software called SPSS (version 21th). In this was applied the classification technique by hierarchical clusters in a training set of 500 cases. This gave the result of the classification for groups of 3, 4 and 5 clusters. This was obtained using the following methods: intergrupos, intragrupos, vecino más cercano, vecino más lejano, medians aggrupation and Ward method.

To identify the best segmentation, it was taken into account in the classification techniques 3 aspects: an error classified as “cuadrático medio bajo”, the most

appropriated data distribution and the features most relevant of the selected method. These features were presented to the group with 3 clusters into Ward method.

Finally it was developed a characterization of cases grouped in each of the three clusters. In where was found that the variables which are most associated with dimension number 1, are type of affiliated member (PO), type of business and IBC. While the variables which are associated with the dimension number 2, are: age, gender and the region (district).

## 1. INTRODUCCIÓN

Dada la importancia para un fondo de pensiones obligatorias de conocer mejor a sus clientes, identificar sus necesidades para colocar éstas en el centro de sus decisiones y apuntarle a la posibilidad de personalizar sus servicios con el fin de ofrecer un mejor producto a sus afiliados, permitiendo un mejor crecimiento del fondo de pensiones obligatorias y un mejor posicionamiento en el mercado. En este orden de ideas la segmentación de mercados ofrece esta posibilidad, ya que permite formar grupos de clientes con características similares a partir de variables sociodemográficas, de empleador y de origen cliente. Para este trabajo se realizó una segmentación utilizando una técnica estadística computacional por conglomerados jerárquicos mediante la metodología de investigación Knowledge Discovery in Datamining (KDD) con el fin de conocer información potencialmente útil dentro de una base datos de los afiliados a un fondo de pensiones obligatorias privado en Colombia.

Para realizar el procesamiento de la información se utilizó el software SPSS (versión 21), con el propósito de conocer la existencia de grupos homogéneos que puedan existir en la base de datos. Se procesaron 500 casos de entrenamiento que contienen variables cualitativas, y se obtuvo el resultado de la clasificación para 3, 4 y 5 conglomerados utilizando los métodos vecino más cercano, vecino más lejano, vinculación intragrupos, vinculación intergrupos, agrupación de centroides, agrupación de mediana y método de Ward.

Posteriormente se identifica el grupo con el mejor número de conglomerados; éste se realizó comparando el error cuadrático medio entre las técnicas de clasificación, la mejor distribución de los datos de acuerdo al criterio de un experto y las buenas características del método seleccionado según la teoría. De acuerdo a lo anterior el grupo con 3 conglomerados presentó las mejores características utilizando el método de Ward, luego se realizó la interpretación de los afiliados segmentados en el grupo con 3 conglomerados.

## **2. PROBLEMA**

### **2.1 PLANTEAMIENTO DEL PROBLEMA**

Un fondo de pensiones obligatorias privado en Colombia no tiene información segmentada que permita identificar y agrupar características en común o similitudes entre sus afiliados. En este sentido el fondo de pensiones obligatorias presenta dificultades para generar estrategias orientadas en la fidelización, retención o recordación de clientes con buenas proyecciones de crecimiento en el medio o largo plazo, generar campañas de ahorro para una vejez digna y optimizar canales de comunicación con el cliente.

Por observación o análisis de un experto es difícil poder identificarlas, por lo que se debe acudir a la aplicación de una técnica estadística computacional que permita el agrupamiento de afiliados con características cualitativas similares que clasifique los afiliados para que sean homogéneos en el segmento y heterogéneos en segmentos diferentes.

### **2.2 FORMULACIÓN DEL PROBLEMA**

Por lo anterior para el presente estudio se formula la siguiente pregunta de investigación, ¿cuáles son las características de los conglomerados obtenidos de la implementación de una técnica estadística computacional de clasificación, para segmentar los afiliados a un fondo de pensiones obligatorias?

### **3. OBJETIVO**

#### **3.1 OBJETIVO GENERAL**

Segmentar los afiliados a un fondo de pensiones obligatorias a partir de una técnica estadística computacional de clasificación por conglomerados jerárquicos.

#### **3.2 OBJETIVOS ESPECÍFICOS**

- Realizar un análisis descriptivo de las variables edad, ingreso base de cotización, género, tipo de afiliado, regional y tipo de empresa mediante la aplicación de un análisis de correspondencia múltiple.
- Aplicar una técnica estadística computacional de clasificación por conglomerados jerárquicos que permita segmentar los afiliados a un fondo de pensiones obligatorias.
- Interpretar los resultados de la segmentación a los afiliados de un fondo de pensiones obligatorias mediante la aplicación de una técnica estadística computacional de clasificación por conglomerados jerárquicos.



#### **4. JUSTIFICACIÓN**

La importancia del siguiente trabajo surge de la necesidad de implementar una técnica estadística computacional por conglomerados jerárquicos, que permita segmentar los afiliados con características en común o similitudes a un fondo de pensiones obligatorias que tiene un volumen considerable de datos y que por observación o análisis de un experto es difícil identificar esas características. Aplicando la técnica estadística computacional la AFP puede tener herramientas para generar estrategias de fidelización, retención o recordación de clientes con buenas proyecciones de crecimiento en el medio o largo plazo, generar campañas de ahorro para una vejez digna, optimizar canales de comunicación con el cliente, direccionar esfuerzos y recursos a grupos de afiliados con ciertas características, como por ejemplo gestión proactiva de retención, gestión comercial de satisfacción, brochure informativo de ventajas, prioridad PQRS, entre otras.

Los resultados de este trabajo permitieron evidenciar las ventajas de implementar la metodología KDD con el fin de conocer información potencialmente útil dentro de los datos utilizando la técnica estadística computacional de clasificación por conglomerados jerárquicos, en razón a que se puede realizar satisfactoriamente un análisis descriptivo multivariado a la hora de segmentar los afiliados en un fondo de pensiones obligatorias.

## **5. MARCO TEÓRICO**

### **5.1 Segmentación de mercado: estrategias básicas para identificar segmentos y elegir un mercado meta**

Con poblaciones y comunidades cada día más diversas, sería arriesgado para una firma en particular ofrecer la misma estrategia comercial a tan distintos consumidores. La segmentación de mercado ofrece a las empresas la posibilidad de personalizar o confeccionar a la medida una combinación única de elementos como el producto, el precio, la plaza y la promoción, para mercados meta específicos. Por consiguiente, les permite satisfacer las necesidades de sus clientes de una manera más efectiva, a través de una propuesta de valor potencialmente superior a la de la competencia. El término segmentación de mercado se refiere entonces al proceso de definición y división de un amplio mercado en grupos claramente identificables y homogéneos de consumidores con necesidades, deseos y características similares. Se dice que un segmento de mercado debería ser:

- Fácil y claramente identificable
- Medible
- Accesible a través de promoción, comunicación y canales de distribución
- Diferente en su reacción a determinada propuesta de valor
- Durable (que no cambie tan rápidamente)
- Apropiado para las políticas y recursos de la empresa
- Suficientemente grande para ser rentable

La base de datos para la segmentación es un factor que varía entre los grupos de un mercado, pero que es consistente al interior de los propios grupos. Todo mercado puede ser segmentado de varias maneras, y aunque muchas de las

bases utilizadas para segmentar un mercado orientado al consumidor también se pueden aplicar a los negocios y a la industria, la naturaleza de estos eventualmente da origen a otras bases específicas de segmentación. En seguida se muestra una tabla comparativa de algunos criterios o estrategias básicas para segmentar ambos tipos de mercados<sup>1</sup>:

**Tabla 1. Mercado Orientado al Consumidor y a la Industria**

<b>Mercado orientado al consumidor</b>		<b>Mercado orientado a la industria</b>	
<b>Tipo de segmentación</b>	<b>Variables consideradas</b>	<b>Tipo de segmentación</b>	<b>Variables consideradas</b>
Geográfica	Región, clima, densidad y tasa de crecimiento de la población	Geográfica	Ubicación, concentración de clientes, tasa de crecimiento de la industria local y factores macroeconómicos diversos
Demográfica	Edad, género, origen étnico, nacionalidad, educación, ocupación, religión, ingreso y estatus familiar	Por tipo de cliente	Tamaño de la organización, industria del cliente y posición en la cadena de valor
Psicográfica	Valores, actitudes, opiniones, intereses, actividades y estilos de vida de la población	Por comportamiento	Lealtad hacia los proveedores, patrones de utilización y tamaño de las órdenes
Por comportamiento	Nivel y patrones de utilización de productos, sensibilidad de precio, lealtad hacia una marca y búsqueda de beneficios y ofertas		

## 5.2 Análisis de Correspondencias (ACM)

Es una técnica descriptiva o exploratoria cuyo objetivo es resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible. En esta línea, su objetivo es similar al de los métodos factoriales, salvo que en el caso del análisis de correspondencias el método se aplica sobre variables categóricas. El análisis de *correspondencias simples* se utiliza a menudo en la representación de datos que se pueden presentar en forma

<sup>1</sup> <http://tisconsulting.org/es/news/market-segmentation-basic-strategies/>

de tablas de contingencia de dos variables nominales u ordinales. Otras utilizaciones implican el tratamiento de tablas de proximidad o distancia entre elementos y tablas de preferencias.

Si se trata de una tabla de contingencia de dos variables cualitativas, una variable cuyas categorías aparecen en filas y la otra variable cuyas categorías son representadas en columnas, el análisis de correspondencias consiste en resumir la información presente en las filas y columnas de manera que pueda proyectarse sobre un subespacio reducido, y representarse simultáneamente los puntos fila y los puntos columna, pudiéndose obtener conclusiones sobre relaciones entre las dos variables nominales u ordinales de origen.

La extensión del análisis de correspondencias simples al caso de varias variables categóricas (tablas de contingencia multidimensionales) se denomina *Análisis de Correspondencias Múltiples*, y utiliza los mismos principios generales que la técnica anterior. En general se orienta a casos en los cuales una variable representa ítems o individuos y el resto son variables cualitativas u ordinales que representan cualidades.

Entre la utilización del Análisis de Correspondencias Simple y Múltiple, se tienen los siguientes estudios:

- ✓ Preferencias de consumo en investigación de mercados.
- ✓ Posicionamiento de empresas a partir de las preferencias de consumidores.
- ✓ Búsqueda de tipologías de individuos respecto a variables cualitativas (patrones de enfermedades en medicina, perfiles psicológicos, comportamiento de especies en biología, etc.).<sup>2</sup>

---

<sup>2</sup> <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>

### **5.3 Análisis de Conglomerados Jerárquicos**

El proceso de conglomerados jerárquicos permite aglomerar tanto casos como variables entre una gran variedad de métodos de aglomeración y medidas de distancia.

El análisis de conglomerados jerárquicos comienza con el cálculo de la matriz de distancias entre los elementos de la muestra (casos o variables). Esa matriz contiene las distancias existentes entre cada elemento y todos los restantes de la muestra. Posteriormente se buscan los elementos más próximos (es decir, los dos más similares en términos de distancia) y se agrupan en un conglomerado. El conglomerado resultante es indivisible a partir de ese momento: de ahí el nombre de jerárquico asignado al procedimiento. De esta manera se van agrupando los elementos en conglomerados cada vez más grandes y heterogéneos hasta llegar al último paso, en el que todos los elementos muestrales quedan agrupados en único conglomerado global. En cada paso del proceso pueden agruparse casos individuales, conglomerados previamente formados o un caso individual con un conglomerado previamente formado. El análisis de conglomerados jerárquico es, por tanto, una técnica aglomerativa: partiendo de los elementos muestrales individualmente considerados, va creando grupos hasta llegar a la formación de un único grupo o conglomerado constituido por todos los elementos de la muestra.

El procedimiento de conglomerados jerárquicos del SPSS informa de todos los pasos realizados en el análisis, por lo que resulta fácil apreciar que elementos o conglomerados se han fusionado en cada paso y a qué distancia se encontraban cuando se han fusionado. Esto permite valorar la heterogeneidad de los conglomerados que se van fundiendo en cada etapa del análisis y decidir en cuál de ellas la fusión de elementos incrementa excesivamente la heterogeneidad de los conglomerados. Aunque el análisis termina cuando se ha terminado de agrupar todos los casos en único conglomerado, el objetivo del analista será el de

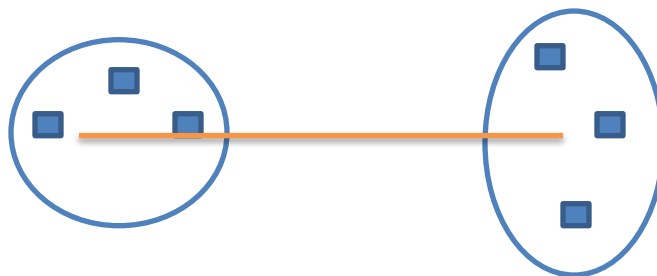
descubrir la existencia de grupos homogéneos “naturales” que puedan existir en el archivo de datos.

La versatilidad del análisis de conglomerados jerárquicos radica en la posibilidad de utilizar distintos tipos de medidas para estimar la distancia existente entre los casos o las variables, la posibilidad de transformar la métrica original de las variables y la posibilidad de seleccionar de entre una gran cantidad de métodos de aglomeración.<sup>3</sup>

#### 5.4 Tipos de Métodos de Clasificación por Conglomerados Jerárquicos

**Vecino más lejano:** va a asegurar que la distancia máxima dentro de un clúster será menor que la distancia entre cualquiera de sus elementos y los elementos más alejados de los demás clúster, útil para detectar outliers, tiende a construir clúster pequeños y compactos, alargar mucho el proceso y dar como resultado agrupaciones encadenadas, no es útil para resumir datos y construye conglomerados más esféricos.

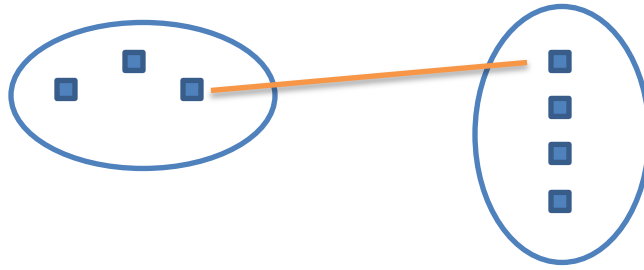
Figura 1. Vecino más Lejano



**Vecino más cercano:** agrupa los casos que se encuentran a menor distancia, unidos dos casos, a continuación se forma el tercer conglomerado buscando la distancia más corta entre los tres elementos, detecta muy bien “outliers”, inconvenientes de encadenamiento y construye conglomerados más alargados.

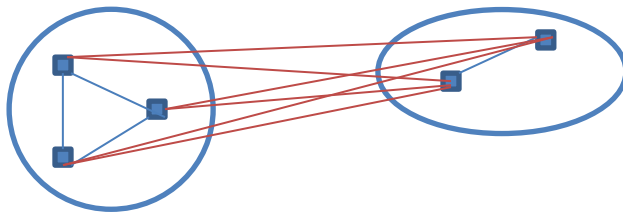
<sup>3</sup> <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/GuiaSPSS/22conglj.pdf>

Figura 2. Vecino más Cercano



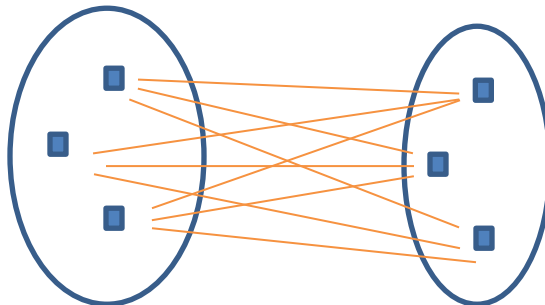
**Vinculación Intergrupos:** la distancia entre los grupos es la media aritmética de las distancias existentes entre todos los componentes de cada grupo, considerados dos a dos, se consiguen grupos con varianzas similares y pequeñas, es sensible a transformaciones monótonas de los datos.

Figura 3. Vinculación Intergrupos



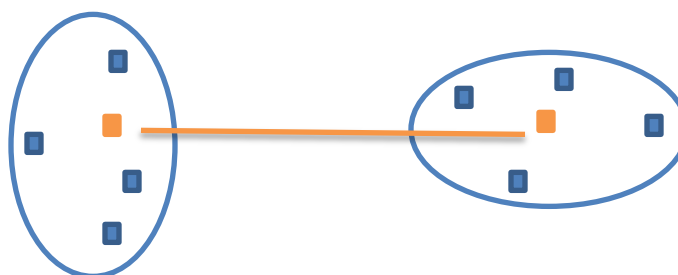
**Vinculación Intragrupos:** se combinan los grupos buscando que la distancia promedio dentro de cada conglomerado sea la menor posible. Así en lugar de considerar los pares de los elementos que pertenecen a cada uno de los grupos, se consideran todos los pares resultantes en caso de que los dos grupos se uniesen.

Figura 4. Vinculación Intragrupos



**Centroide:** el proceso comienza calculando el centro de gravedad de cada conglomerado, para agrupar los conglomerados cuya distancia entre centroides sea mínima. Tras unir dos conglomerados se determinó el nuevo centro de gravedad y se procede de forma similar. Si se fusionan dos grupos de diferente tamaño, el centroide del nuevo grupo queda más cerca del grupo de mayor tamaño y más alejado del de menor tamaño en proporción a sus diferencias de tamaño, construye conglomerados más robustos.

Figura 5. Método Centroide



**Mediana:** la distancia entre los grupos se obtiene calculando la distancia promedio entre todos los pares de observaciones independientemente de que estén próximos o alejados. Agrupa los conglomerados con un tamaño óptimo y fusiona clúster con varianzas pequeñas.<sup>4</sup>

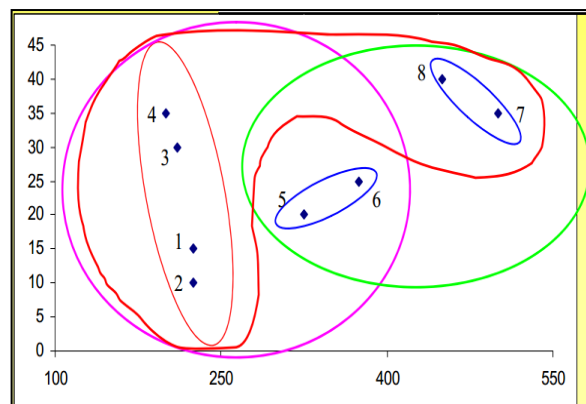
**Método de Ward:** es el más utilizado ya que maximiza la homogeneidad dentro de los grupos. Para ello, plantea todas las posibles fusiones en cada etapa concreta (ver figura No. 6) y elige la que maximiza la homogeneidad, calculando los centroides de los grupos resultantes de las posibles fusiones, luego calcula la distancia al centroide de todas las observaciones (suma de cuadrados total) y por último la menor suma de cuadrados total es la elegida.<sup>5</sup>

<sup>4</sup> *Cómo aplicar un clúster jerárquico en SPSS - Ruth Vilá-Baños, María-José Rubio-Hurtado, Vanesa Berlanga-Silvente, Mercedes Torrado- Fonseca.*

<sup>5</sup> *Universidad de Córdoba, Análisis clúster, Prof. Dr. José Perea, Dpto. producción.*



Figura 6. Método de Ward



## 5.5 Identificación de Variables Objeto de Estudio

La siguiente tabla muestra las 6 variables con el score más alto según un estudio de investigación que realizó el fondo de pensiones obligatorias.

Tabla 2. Identificación de Variables Objeto de Estudio

Tipo Variable	P/C	Variable	Descripción	Score
INFORMACION SOCIODEMOGRAFICA CLIENTE	P42	Edad	16 y 26 años	0,7868
	P43		27 y 37 años	0,3034
	P44		38 y 49 años	0,4667
	C45		50 o más años	0,1584
			<b>Total</b>	<b>1,7153</b>
	P48	Regional	Centro	0,9333
	P49		Antioquia	0,2579
	P50		Occidente	0,2713
	P51		Norte	0,3945
			<b>Total</b>	<b>1,857</b>
	C53	IBC	4.284.801 y 8.569.600	1,21
	C54		2.142.401 y 4.284.800	0,685
	C55		1.071.201 y 2.142.400	0,4887
	C56		535.600 y 1.071.200	0,3096
			<b>Total</b>	<b>2,6933</b>
	P33	Tiene Beneficiario	Si	0,253
	P34		No	-0,18
			<b>Total</b>	<b>0,073</b>
INFORMACION EMPLEADOR	P24	Naturaleza Aportante	Persona Natural	0,175
	P25		Persona Juridica	0,234
			<b>Total</b>	<b>0,409</b>
	C51	Tamaño	Corporativo	0,154
	C52		Micro	0,0745
	C53		Pyme	0,234
			<b>Total</b>	<b>0,4625</b>
	P48	Tipo de empresa	Superior	0,1913
	P49		Alto	0,234
	P50		Medio	0,567
	P51		Bajo	0,765
			<b>Total</b>	<b>1,7573</b>
	C38	Antigüedad relación laboral	De 1 a 19 meses	0,189
	C39		De 20 a 48 meses	0,1946
	C40		De 49 a 60 meses	0,126
	C41		De 61 a 84 meses	0,127
	C42		De 85 a 120 meses	0,256
	C43		Mayor a 120 meses	0,365
			<b>Total</b>	<b>1,2576</b>
ORIGEN CLIENTE	C58	Genero	M	0,7357
	C59		F	0,896
			<b>Total</b>	<b>1,6317</b>
	P2	Tipo de afiliado PO	Dependiente	0,507
	P3		Independiente	0,495
	P4		Mixto	0,671
			<b>Total</b>	<b>1,673</b>
	C4	Canal de venta PO	Corredor	0,117
	C5		Ejecutivo	0,158
	C6		Asesor	0,654
			<b>Total</b>	<b>0,929</b>
	C11	Origen PO	Traslado de régimen	0,496
	C12		Traslado de AFP	0,267
	C13		Afiliado Nuevo	0,156
			<b>Total</b>	<b>0,919</b>

## 6. MARCO LEGAL

La AFP es una organización de servicios de ahorro previsional confiable, orientada al cliente, que cumple con los servicios de pensión de un trabajador, ofreciendo rentabilidad y flexibilidad, basándose en la ética, excelencia y profesionalismo, que está reglamentada con la ley 100 que obliga tanto al trabajador como al empleador a realizar aportes en dinero durante toda la vida laboral del empleado, consolidando en un futuro una vejez digna y sin preocupaciones económicas para el empleado y su familia. Cuando un trabajador es afiliado a Pensión tiene derecho a<sup>6</sup>:

- ✓ Pensión de Vejez: es el pago a que tiene derecho el trabajador una vez se retire de su vida laboral por edad para garantizar una vejez digna.
- ✓ Pensión de Invalidez: es el pago a que tiene derecho el trabajador por la pérdida o disminución de su capacidad laboral.
- ✓ Pensión de Sobrevivencia: es el pago a que tienen derecho los miembros del núcleo familiar por muerte del afiliado o pensionado.

Tabla 3. Normas que Rigen la Seguridad Social en Colombia

LEY	DESCRIPCIÓN
100 de 1993	El sistema de seguridad social integral es el conjunto de instituciones, normas y procedimientos, de que disponen la persona y la comunidad para gozar de una calidad de vida, mediante el cumplimiento progresivo de los planes y programas que el Estado y la sociedad desarrollen para proporcionar la cobertura integral de las contingencias, especialmente las que menoscaban la salud y la capacidad económica, de los habitantes del territorio nacional, con el fin de lograr el bienestar individual y la integración de la comunidad.
797 de 2003	Por la cual se reforman algunas disposiciones del sistema general de pensiones previsto en la Ley 100 de 1993 y se adoptan disposiciones sobre los Regímenes Pensionales exceptuados y especiales.

<sup>6</sup> (Dirección Nacional de Capacitación, 2005)

## 7. MARCO METODOLÓGICO

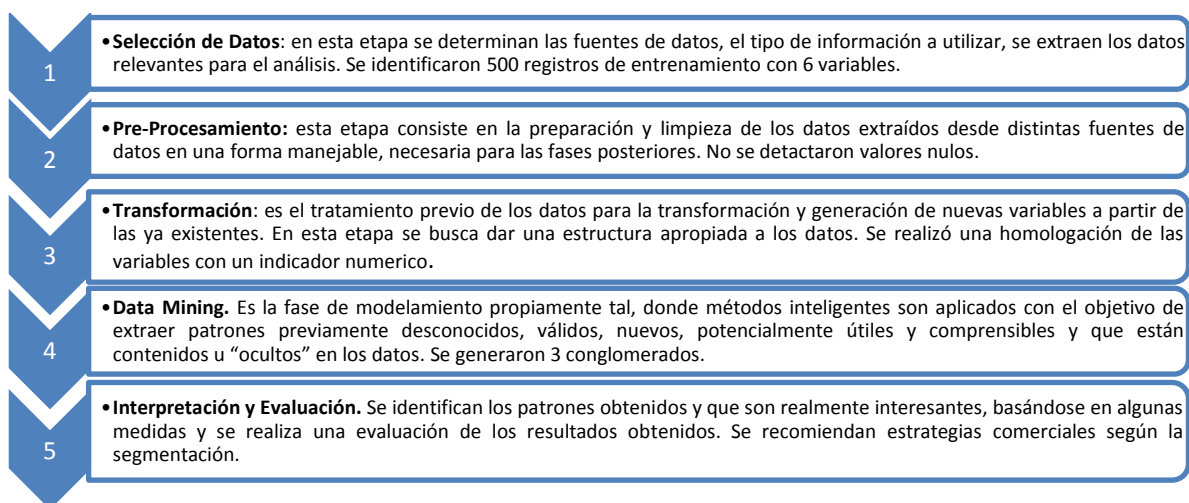
### 7.1 Datos de Estudio

La base de datos utilizada en el presente trabajo está conformada por una población de 6,381 afiliados vinculados en diciembre de 2011 a un fondo privado de pensiones obligatorias, las cuales se encontraban activas al momento del estudio y laboraban en distintas regiones del país, la base fue particionada para obtener un conjunto de datos de entrenamiento de 500 afiliados con variables relevantes para el estudio como son: edad, regional, ingreso base de cotización (IBC), genero, tipo de afiliación y tipo de empresa.

### 7.2 Tipo de Investigación – Análisis Descriptivo

En este trabajo se empleó la metodología KDD (Knowledge Discovery in Databases), que se basa en la búsqueda de información relevante dentro de una base de datos con gran volumen de información. Lo anterior con el fin de establecer relaciones o modelos de los datos para la oportuna toma de decisiones. Esta metodología está conformada por las siguientes etapas<sup>7</sup>:

Figura 7. Metodología KDD



<sup>7</sup> <http://www.webmining.cl/2011/01/proceso-de-extraccion-de-conocimiento/>

### 7.3 Variables de Estudio

La siguiente tabla muestra las variables y su estructura que serán utilizadas en nuestro estudio de investigación:

**Tabla 4. Variables Cualitativas**

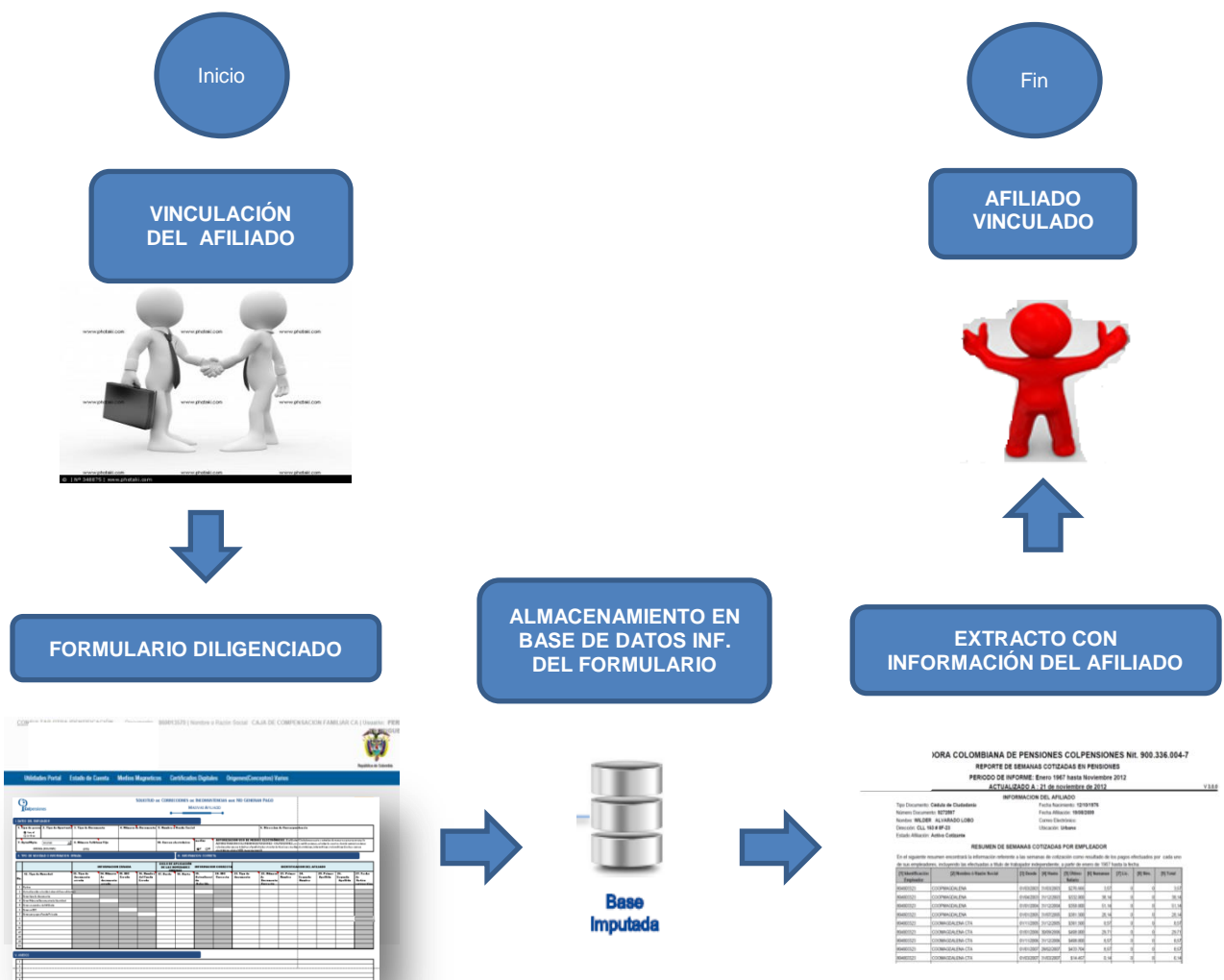
<b>NOMBRE DE LA VARIABLE</b>	<b>DEFINICIÓN</b>	<b>DESCRIPCIÓN</b>
Edad	Está definida por rangos de edad: 1) 16 y 26 años 2) 27 y 37 años 3) 38 y 49 años 4) 50 o más años	Los rangos están definidos por la fecha de nacimiento de los afiliados al fondo de pensiones obligatorias.
Regional	Está definida por Regional: 1) Centro 2) Antioquia 3) Occidente 4) Norte	La regional corresponde a la ciudad de vinculación del afiliado.
IBC	Está definida por rangos de IBC: 1) 4.284.801 y 8.569.600 2) 2.142.401 y 4.284.800 3) 1.071.201 y 2.142.400 4) 535.600 y 1.071.200	Los rangos están definidos por el IBC reportado al momento de realizar la vinculación.
Genero	Está definida por Genero: 1) Masculino 2) Femenino	El género fue diligenciado en el formato de vinculación.
Tipo de afiliado PO	Está definida por tipo de afiliado: 1) Dependiente 2) Independiente	Dependiente: presenta un contrato laboral con una empresa. Independiente: El afiliado realizó sus aportes a seguridad social.
Tipo de empresa	Está definida por tipo de empresa: 1) Superior 2) Alto 3) Medio 4) Bajo	La calificación está determinada por las ventas y el número de trabajadores de la empresa donde labora el afiliado.

Fuente: Elaboración basada en estudios de la empresa.

## 7.4 Recolección de Información

La información relacionada a las variables consideradas en el presente estudio, fue obtenida a partir de los datos registrados en el sistema de información AS400, que son cargados cuando se realiza una vinculación de una persona al fondo de pensiones obligatorias. Para este estudio se utilizará la información recolectada en el módulo de Pensiones Obligatorias. Existen unas tablas socio demográficas, de empleador y de cliente donde se almacena la información relacionada con la fecha de nacimiento, género, tipo de afiliado, regional, fecha de afiliación e Ingreso Base de Cotización (IBC) y otra tabla donde se encuentra el tipo de empresa. Las fuentes que alimentan estas tablas dependen de la relación de la AFP con sus clientes (afiliados o empleadores) y mensualmente se van almacenando con procesos manuales de digitación de los respectivos formatos de vinculación.

Figura 8. Procedimiento de Recolección de Información



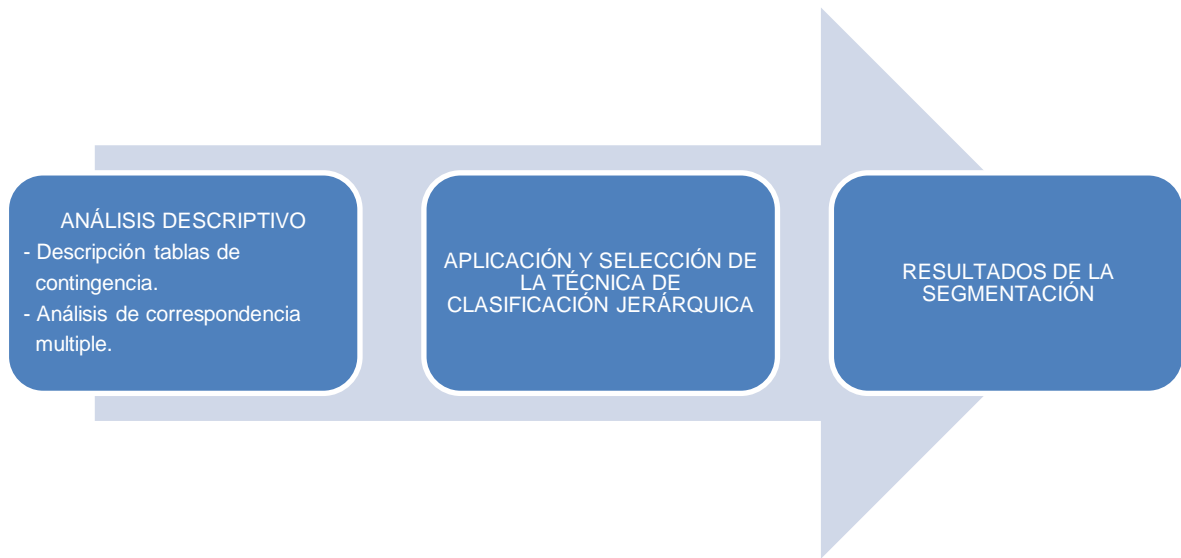
Una vez almacenada la información en la plataforma tecnológica de la empresa AS400, se realiza el proceso de extracción de información de pensiones obligatorias, que proporciona una base de datos de 6.381 registros organizados por fecha de vinculación y que presenta las siguientes variables: fecha de nacimiento (transformada en edad), género, regional, tipo de afiliado (PO), tipo de empresa e IBC.

Una vez definidas las variables a estudiar se continuó con el proceso de generación de una base de entrenamiento con los primeros 500 registros organizados por fecha de afiliación, a partir del particionamiento de la base de 6.381 registros, con esta prueba piloto se desea conocer como es el comportamiento, asociación, clasificación y segmento de los datos, mediante la implementación de la técnica de clasificación por conglomerados jerárquicos a partir de la estadística computacional. Posteriormente se realizó la detección de outliers que fue de 0 en la base de entrenamiento. Las variables como edad e IBC fueron transformadas por rangos previo estudio del fondo de pensiones obligatorias y luego se etiquetaron todas las variables con un número específico (ver tabla 4).

## 7.5 Diseño Estadístico

El tratamiento estadístico que se le dio a la información analizada o estudiada en el presente trabajo constó de las siguientes etapas:

**Figura 9. Proceso Estadístico**



Una vez depurada la base de datos y categorizadas las variables, se procede a hacer un análisis descriptivo de variables cualitativas, como género, regional, IBC, tipo de afiliado, edad y tipo de empresa, para conocer el número de casos en cada una de las categorías relacionadas, lo anterior haciendo uso de tablas de contingencia.

La identificación del método de clasificación jerárquica y el número de conglomerados más eficiente, dependerá de la comparación del error cuadrático medio entre las técnicas de clasificación, la mejor distribución de los datos de acuerdo a los criterios de un experto y las buenas características del método seleccionado según la teoría.



## 8. RESULTADOS

### 8.1 Análisis Descriptivo

#### 8.1.1 Descripción Tablas de Contingencia

Inicialmente se realizó una descripción de los 500 afiliados al fondo de pensiones obligatorias de diciembre de 2011 considerados en el estudio, teniendo en cuenta las variables, IBC, tipo de empresa, edad y regional. Los resultados relacionados se presentan en la Tabla 5. En esta se observan 10 afiliados que tienen IBC entre 4.284.801 y 8.569.600, con edades entre 38 y 49 años, tipo de empresa superior y que están en la regional Centro. Con el mismo IBC se observan 9 afiliados con edades entre 27 y 37 años, tipo de empresa superior y que están en la regional Centro. En esta tabla se observan 21 afiliados que tienen IBC entre 2.142.401 y 4.284.800, con edades entre 27 y 37 años, tipo de empresa superior y que están en la regional Centro. Con el mismo IBC se observan 6 afiliados con edades entre 38 y 49 años, tipo de empresa superior y que están en la regional Centro.

**Tabla 5. Estadísticos Descriptivos IBC entre 4.284.801 y 8.569.600 – IBC entre 2.142.401 y 4.284.800**

		IBC							
		IBC entre 4.284.801 y 8.569.600				IBC entre 2.142.401 y 4.284.800			
		Tipo de empresa				Tipo de empresa			
		Superior	Alto	Medio	Bajo	Superior	Alto	Medio	Bajo
Región	Edad	No.	No.	No.	No.	No.	No.	No.	No.
Centro	16 y 26	0	0	0	0	0	0	0	0
	27 y 37	9	0	1	3	21	0	1	1
	38 y 49	10	3	4	0	6	1	1	2
	>= 50	2	0	0	0	1	0	0	0
Antioquia	16 y 26	0	0	0	0	0	0	0	0
	27 y 37	0	0	0	0	1	0	2	0
	38 y 49	0	0	1	0	0	0	3	0
	>= 50	0	0	0	0	0	0	0	0
Occidente	16 y 26	0	0	0	0	0	0	0	0
	27 y 37	0	0	0	1	0	0	1	1
	38 y 49	0	0	0	0	1	0	1	0
	>= 50	0	0	0	0	0	0	0	0
Norte	16 y 26	0	0	0	0	0	0	0	0
	27 y 37	0	0	0	0	1	1	1	0
	38 y 49	2	0	1	0	1	0	0	0
	>= 50	0	0	0	0	0	0	0	0

En la tabla 6 se observan 14 afiliados que tienen IBC entre 1.071.201 y 2.142.400, con edades entre 27 y 37 años, tipo de empresa superior y que están en la regional Centro. En esta tabla se observan 63 afiliados que tienen IBC entre 535.600 y 1.071.200, con edades entre 27 y 37 años, tipo de empresa alto y que están en la regional Norte. Con el mismo IBC, edad y tipo de empresa se observan 43 afiliados que están en la Regional Centro y 25 afiliados que están en la Regional Antioquía.

**Tabla 6. Estadísticos Descriptivos IBC entre 1.071.201 y 2.142.400 – IBC entre 535.600 y 1.071.200**

Región	Edad	IBC							
		IBC entre 1.071.201 y 2.142.400				IBC entre 535.600 y 1.071.200			
		Tipo de empresa				Tipo de empresa			
		Superior	Alto	Medio	Bajo	Superior	Alto	Medio	Bajo
		No.	No.	No.	No.	No.	No.	No.	No.
Centro	16 y 26	0	0	0	0	0	2	1	1
	27 y 37	14	3	3	0	7	43	1	8
	38 y 49	3	0	1	0	10	15	0	3
	>= 50	0	0	0	0	1	0	0	0
Antioquia	16 y 26	0	0	0	0	0	2	0	1
	27 y 37	2	1	1	0	11	25	1	6
	38 y 49	0	0	0	0	6	12	0	0
	>= 50	0	0	0	0	0	3	0	0
Occidente	16 y 26	0	0	0	0	0	0	0	0
	27 y 37	3	1	0	2	7	16	0	11
	38 y 49	1	0	1	0	6	12	0	5
	>= 50	0	0	0	0	0	1	0	0
Norte	16 y 26	0	0	0	0	0	1	0	0
	27 y 37	3	1	4	1	12	63	0	16
	38 y 49	3	2	0	0	7	33	0	8
	>= 50	0	1	0	0	0	7	0	0

### 8.1.2 Análisis de Correspondencia Múltiple (ACM)

A continuación se presenta el análisis de correspondencias múltiples de los 500 casos de entrenamiento, teniendo en cuenta las variables edad, género, tipo de afiliado, tipo de empresa, regional e IBC.

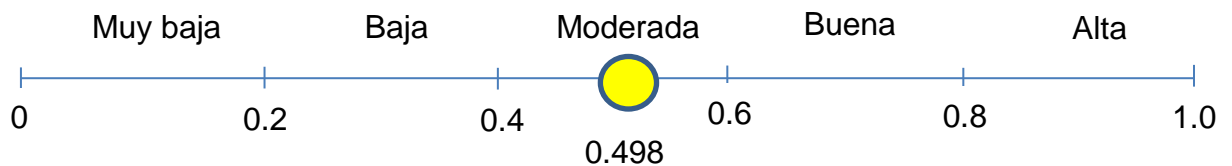
**Tabla 7. Resumen del Modelo**

Resumen del modelo				
Dimensión	Alfa de Cronbach	Varianza explicada		
		Total (Autovalores)	Inercia	% de la varianza
1	,573	1,915	,319	31,912
2	,402	1,504	,251	25,071
Total		3,419	,570	
Media	,498 <sup>a</sup>	1,709	,285	28,491

a. El Alfa de Cronbach Promedio está basado en los autovalores promedio.

El Alfa de Cronbach mide la asociación de las variables en las dimensiones mediante la consistencia interna, en nuestro caso el indicador corresponde al 0.498, al incluir variables como género y tipo de afiliado el Alfa de Cronbach baja al 0.474. Según la gráfica el modelo presenta una consistencia interna moderada.

**Figura 10. Escala de consistencia interna – Alfa de Cronbach**



La dimensión 1 explica el 31,9% de la variabilidad del modelo, mientras que la dimensión 2 explica el 25% de la variabilidad del modelo, lo que indica que con 2 dimensiones se explica el 57% de la solución global.

**Tabla 8. Medidas de Discriminación**

	Dimensión		Media
	1	2	
Edad	,028	,090	,059
Genero	,004	,514	,259
Regional	,376	,562	,469
Tipo de afiliado PO	,080	,061	,071
Tipo de empresa	,679	,223	,451
IBC	,747	,054	,400
Total	1,915	1,504	1,709
% de la varianza	31,912	25,071	28,491

De acuerdo a las medidas de discriminación presentadas en la Tabla 8, se puede apreciar que las variables que más se asocian a la dimensión 1 son tipo de afiliado PO (0,080), tipo de empresa (0,679), IBC (0,747), mientras que para la dimensión 2, las variables que más se asocian a ésta son: edad (0.090), género (0.514), regional (0,562), de acuerdo a la puntuación más alta en cada dimensión.

**Figura 11. Medidas de Discriminación**

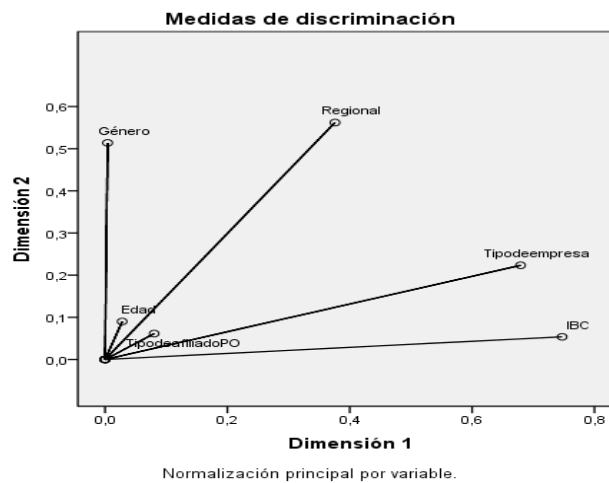
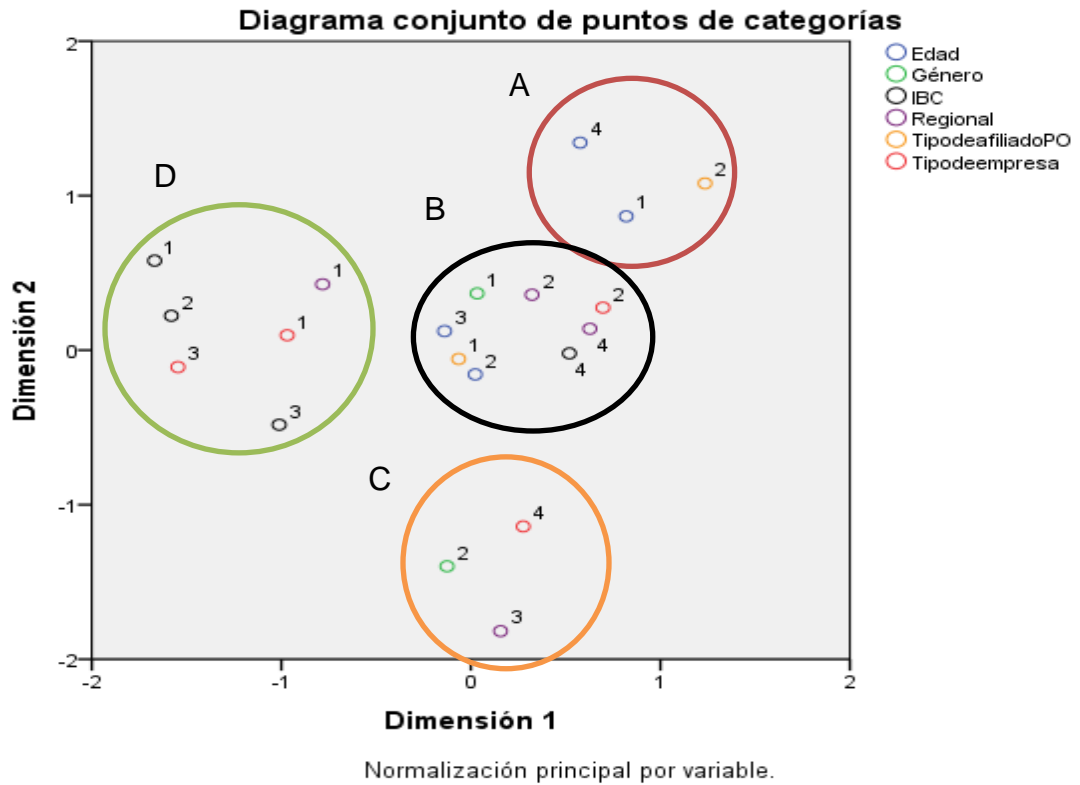


Figura 12. Diagrama Conjunto de Puntos de Categorías



De acuerdo al diagrama conjunto de puntos de categorías se realizó esta asociación de variables, conformadas de la siguiente manera: los afiliados independientes dentro del círculo de color rojo se asocian con afiliados que tienen edades entre 16 y 26 años y mayores a 50 años. Los afiliados dentro del círculo negro, presentan fuerte asociación con los hombres vinculados en la Regional Antioquia y Norte, tipo de empresa alto, tipo de afiliación PO dependiente, edades que oscilan entre los 27 y los 49 años e ingresos entre 535.600 y 1.071.200. Los afiliados dentro del círculo naranja, presentan fuerte asociación dada por el género Femenino, regional occidente y tipo de empresa bajo. Los afiliados dentro del círculo verde, se relacionan por tener ingresos mayores a 1.071.201, vinculados en la Regional Centro y que corresponden al tipo de empresa superior y medio.

Tabla 9. Descripción de Variables Según Diagrama Conjunto de Puntos de Categoría

Clasificación	Identificador	Variable	Tipo de variable
A	1	Edad	16 y 26
	4		> = 50
	2	Tipo de afiliado PO	Independiente
B	1	Género	Masculino
	2	Edad	27 y 37
	3		38 y 49
	1	Tipo de afiliado PO	Dependiente
	4	IBC	535.600 Y 1.071.200
	2	Tipo de empresa	Alto
	2	Regional	Antioquía
	4		Norte
C	4	Tipo de empresa	Bajo
	2	Género	Femenino
	3	Regional	Occidente
D	1	IBC	4.284.801 Y 8.569.600
	2		2.142.401 Y 4.284.800
	3		1.071.201 Y 2.142.400
	1	Tipo de empresa	Superior
	3		Medio
	1	Regional	Centro

La Tabla 9 sirve de apoyo para entender el diagrama de puntos de categoría, porque describe la estructura de las variables.

## 8.2. Aplicación y Selección de la Técnica de Clasificación Jerárquica

A continuación se presentan los resultados obtenidos luego de aplicar las diferentes técnicas de clasificación por conglomerados jerárquicos, variando para cada clúster el número de conglomerados de 1 a 5.

Tabla 10. Clasificación Jerárquica Grupo con 3 Conglomerados

Conglomerado	Intergrupos	%	Intragrupos	%	Vecino mas cercano	%	Vecino mas lejano	%	Agrupacion de medianas	%	Agrupacion centroides	%	Metodo de ward	%
1	235	47%	246	49%	118	24%	268	54%	322	64%	402	80%	148	30%
2	5	1%	187	37%	381	76%	105	21%	177	35%	5	1%	251	50%
3	260	52%	67	13%	1	0,2%	127	25%	1	0%	93	19%	101	20%
<b>Total general</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>
<b>ECM</b>	<b>0,705574943</b>		<b>0,993730346</b>		<b>0,428070088</b>		<b>0,842897384</b>		<b>0,483565921</b>		<b>0,77979228</b>		<b>0,6994026</b>	
Varianza	0,498833667		0,989247571		0,182799334		0,7118998		0,234517227		0,609294589		0,49014429	

Tabla 11. Clasificación Jerárquica Grupo con 4 Conglomerados

Conglomerado	Intergrupos	%	Intragrupos	%	Vecino mas cercano	%	Vecino mas lejano	%	Agrupacion de medianas	%	Agrupacion centroides	%	Metodo de ward	%
1	235	47%	195	39%	118	23,6%	268	54%	296	59%	402	80%	148	30%
2	5	1%	187	37%	379	75,8%	13	3%	177	35%	5	1%	182	36%
3	160	32%	51	10%	2	0,4%	127	25%	26	5%	72	14%	69	14%
4	100	20%	67	13%	1	0,2%	92	18%	1	0,2%	21	4%	101	20%
<b>Total general</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100,0%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>
ECM	1,013706072		1,235920709		0,442736942		1,23231652		0,603907278		0,885564227		1,0869609	
Varianza	1,029659319		1,5304907		0,19560406		1,52164729		0,365735487		0,785795591		1,1838517	

Tabla 12. Clasificación Jerárquica Grupo con 5 Conglomerados

Conglomerado	Intergrupos	%	Intragrupos	%	Vecino mas cercano	%	Vecino mas lejano	%	Agrupacion de medianas	%	Agrupacion centroides	%	Metodo de ward	%
1	235	47%	195	39%	25	5%	268	54%	296	59%	205	41%	148	30%
2	5	1%	187	37%	379	76%	13	3%	122	24%	5	1%	182	36%
3	160	32%	51	10%	2	0,4%	92	18%	55	11%	197	39%	69	14%
4	73	15%	54	11%	93	19%	92	18%	26	5%	72	14%	45	9%
5	27	5%	13	3%	1	0,2%	35	7%	1	0,2%	21	4%	56	11%
<b>Total general</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>	<b>500</b>	<b>100%</b>
ECM	1,0760874		1,3295052		0,8401048		1,4265076		0,8886034		1,2647514		1,2937681	
Varianza	1,1602846		1,7712614		0,7050406		2,0390020		0,7919936		1,6028016		1,6771904	

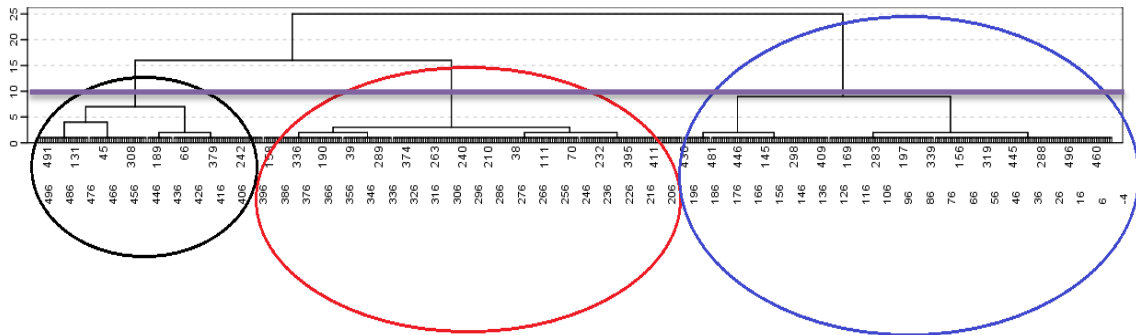
La selección del método de clasificación por conglomerados jerárquicos se realizó a partir de tres factores con un total de 3 conglomerados a conformar, el primero con la observación del (ECM), que correspondió al método Ward (0,69940), el segundo factor se apoyó en la distribución más eficiente al considerar una mayor concentración de individuos en los grupos y se obtuvo con el método de Ward conglomerado 1 (30%), conglomerado 2 (50%) y para el conglomerado 3 (20%) y el tercer factor concluye teóricamente que el método de Ward es uno de los más eficientes, los conglomerados que maneja suelen ser muy compactos y persigue la minimización de la varianza intragrupal y maximiza la homogeneidad dentro de los grupos<sup>8</sup>.

### 8.3 Dendograma resultante de la Técnica de Clasificación Jerárquica por el Método de Ward

La representación gráfica dendograma ayuda a interpretar el resultado de la segmentación, ilustrando como se van haciendo los agrupamientos etapa a etapa.

<sup>8</sup> <http://www.uoc.edu/in3/emath/docs/Cluster.pdf>

**Figura 13. Dendrograma Método de Ward grupo con 3 conglomerados**



En el dendrograma se visualizan los 3 conglomerados y los individuos con características en común en cada uno de estos.

#### 8.4. Resultados de la Técnica de Clasificación Jerárquica con el Método de Ward

Al utilizar el método Ward se debe encontrar el par de clúster que incrementen el mínimo del total de la varianza del clúster después de mezclarlos. Este incremento es la distancia cuadrada con un peso asignado entre los centros de los clúster, la distancia inicial entre los objetos individuales debe ser proporcional al cuadrado de la Distancia Euclidiana, se persigue la minimización de la varianza intragrupal y maximiza la homogeneidad<sup>9</sup>. Como se mencionó anteriormente el número de afiliados agrupados en los 3 clúster, utilizando el método de clasificación por conglomerados jerárquicos es el siguiente:

**Tabla 13. Conglomerados Jerárquicos (3) para el Método de Ward**

Conglomerado	No. de afiliados	%
1	148	30%
2	251	50%
3	101	20%
<b>Total general</b>	<b>500</b>	<b>100%</b>

<sup>9</sup> <http://www.uoc.edu/in3/emath/docs/Cluster.pdf>



Observamos en la tabla No. 13, que el 50% de los afiliados se encuentran en el conglomerado No. 2, el 30% de los afiliados en el conglomerado No. 1 y el 20% de los afiliados en el conglomerado No. 3, las variables utilizadas en el análisis fueron, edad del afiliado, regional de vinculación, tipo de empresa y el Ingreso Base de Cotización (IBC).

A continuación se presenta la caracterización de los 3 conglomerados obtenidos por el método de Ward.

**Tabla 14. Análisis Conglomerados Jerárquicos (conglomerado 1)**

CONGLOMERADO 1					
No.	%	Edad	Regional	Tipo de empresa	IBC
63	43%	27 y 37	NORTE	Alto	535.600 Y 1.071.200
33	22%	38 y 49	NORTE	Alto	535.600 Y 1.071.200
12	8%	27 y 37	NORTE	Superior	535.600 Y 1.071.200
7	5%	38 y 49	NORTE	Superior	535.600 Y 1.071.201
7	5%	27 y 37	OCCIDENTE	Superior	535.600 Y 1.071.200
7	5%	> = 50	NORTE	Alto	535.600 Y 1.071.200
3	2%	38 y 49	NORTE	Superior	1.071.201 Y 2.142.400
3	2%	27 y 37	NORTE	Superior	1.071.201 Y 2.142.400
3	2%	27 y 37	OCCIDENTE	Superior	1.071.201 Y 2.142.400
2	1%	38 y 49	NORTE	Alto	1.071.201 Y 2.142.400
2	1%	38 y 49	NORTE	Superior	>=4.284.801 Y 8.569.600
1	1%	16 y 26	NORTE	Alto	535.600 Y 1.071.200
1	1%	27 y 37	NORTE	Superior	2.142.401 Y 4.284.800
1	1%	> = 50	NORTE	Alto	1.071.201 Y 2.142.400
1	1%	38 y 49	NORTE	Superior	2.142.401 Y 4.284.800
1	1%	38 y 49	OCCIDENTE	Superior	2.142.401 Y 4.284.800
1	1%	38 y 49	OCCIDENTE	Superior	1.071.201 Y 2.142.400
148	100%				

- El número de afiliados está determinado por los afiliados que presentan similitud en las 4 variables.

De acuerdo a los resultados presentados en la tabla 14 se puede apreciar que el conglomerado 1 permite observar que el 78% de los afiliados resaltado en color naranja, tienen un IBC que está en el rango de 535.600 y 1.071.200, trabajan en empresas con un tipo de empresa alto o superior, fueron vinculados en la Regional

Norte y tienen un rango de edad entre 27 y 49 años, donde los trabajadores vinculados a este tipo de empresas alto o superior (ver tabla 14), pueden mantenerse o proyectarse laboralmente con base en estudios realizados por el fondo de pensiones obligatorias.

**Tabla 15. Análisis Conglomerados Jerárquicos (conglomerado 2)**

CONGLOMERADO 2					
No.	%	Edad	Regional	Segmento	IBC
43	17%	27 y 37	CENTRO	Alto	535.600 Y 1.071.200
25	10%	27 y 37	ANTIOQUIA	Alto	535.600 Y 1.071.200
21	8%	27 y 37	CENTRO	Superior	2.142.401 Y 4.284.800
16	6%	27 y 37	OCCIDENTE	Alto	535.600 Y 1.071.200
15	6%	38 y 49	CENTRO	Alto	535.600 Y 1.071.200
14	6%	27 y 37	CENTRO	Superior	1.071.201 Y 2.142.400
12	5%	38 y 49	ANTIOQUIA	Alto	535.600 Y 1.071.200
12	5%	38 y 49	OCCIDENTE	Alto	535.600 Y 1.071.200
11	4%	27 y 37	ANTIOQUIA	Superior	535.600 Y 1.071.200
10	4%	38 y 49	CENTRO	Superior	535.600 Y 1.071.200
10	4%	38 y 49	CENTRO	Superior	>=4.284.801 Y 8.569.600
9	4%	27 y 37	CENTRO	Superior	>=4.284.801 Y 8.569.600
7	3%	27 y 37	CENTRO	Superior	535.600 Y 1.071.200
6	2%	38 y 49	ANTIOQUIA	Superior	535.600 Y 1.071.200
6	2%	38 y 49	OCCIDENTE	Superior	535.600 Y 1.071.200
6	2%	38 y 49	CENTRO	Superior	2.142.401 Y 4.284.800
3	1%	27 y 37	CENTRO	Medio	1.071.201 Y 2.142.400

- La cantidad de afiliados (No.) está determinada por los afiliados que presentan similitud en las 4 variables.

El número de afiliados en el conglomerado 2, de acuerdo a la tabla 15, permite observar que el 33% de los afiliados resaltados en color naranja, tienen IBC que está en el rango de 535.600 y 1.071.200, trabajan en tipos de empresas Alto o Superior, fueron vinculados en la regional Centro, Antioquia y Occidente y tienen un rango de edad entre 27 y 37 años. Mientras el 16% de los afiliados resaltados en color azul presentan un IBC entre 535.600 y 1.071.200, fueron vinculados en la Regional Centro, Antioquia y Occidente, trabajan en empresas con segmento Alto y está en el rango de edad de 38 y 49 años, los afiliados en color verde 14% presentan un IBC entre 1.071.201 y 4.284.800, pertenecen a la Regional Centro y

está en el rango de edad de 27 y 37 años. Este conglomerado se cataloga en el segmento No. 2 los afiliados tienen buenas proyecciones de crecimiento o de mantenerse en la empresa porque está catalogada en tipo de empresas Alto o Superior, los afiliados de color verde tienen buenos ingresos, pertenecen a una Región donde está la capital y presentan la edad para que se puedan desarrollar profesional y laboralmente.

**Tabla 16. Análisis Conglomerados Jerárquicos (conglomerado 3)**

CONGLOMERADO 3					
No.	%	Edad	Regional	Segmento	IBC
16	6%	27 y 37	NORTE	Bajo	535.600 Y 1.071.200
11	4%	27 y 37	OCCIDENTE	Bajo	535.600 Y 1.071.200
8	3%	38 y 49	NORTE	Bajo	535.600 Y 1.071.200
8	3%	27 y 37	CENTRO	Bajo	535.600 Y 1.071.200
6	2%	27 y 37	ANTIOQUIA	Bajo	535.600 Y 1.071.200
5	2%	38 y 49	OCCIDENTE	Bajo	535.600 Y 1.071.200
4	2%	38 y 49	CENTRO	Medio	>=4.284.801 Y 8.569.600
4	2%	27 y 37	NORTE	Medio	1.071.201 Y 2.142.400
3	1%	38 y 49	CENTRO	Bajo	535.600 Y 1.071.200
3	1%	38 y 49	ANTIOQUIA	Medio	2.142.401 Y 4.284.800
3	1%	38 y 49	CENTRO	Alto	>=4.284.801 Y 8.569.600
3	1%	27 y 37	CENTRO	Bajo	>=4.284.801 Y 8.569.600

- La cantidad de afiliados (No.) está determinada por los afiliados que presentan similitud en las 4 variables.

En el conglomerado 3 se puede visualizar de acuerdo a la tabla 16 en color naranja que el 15% de los afiliados se encuentra en un rango de edad entre 27 y 37 años, fueron vinculados en la Regional Centro, Antioquia y Occidente, trabajan en tipo de empresas Bajo y tienen un IBC entre 535.600 y 1.071.200, el 6% de los afiliados en color verde se encuentran en un rango de edad entre 38 y 49 años, están en la Regional Centro, Norte y Occidente, trabajan en empresas de segmento Bajo y tienen un IBC entre 535.600 y 1.071.200, existe un 3% de afiliados en color azul que presentan un rango de edad entre 38 y 49 años, están en la Regional Centro, trabajan en empresas de segmento Medio, Alto y tienen un IBC entre >=4.284.801 y 8.569.600. Este conglomerado lo catalogamos como el

segmento No. 3 el tipo de empresa es bajo, los ingresos de los afiliados son bajos, por lo que presentan pocas posibilidades de crecer en la compañía, según estudios realizados por el fondo de pensiones obligatorias.

A continuación se presenta la tabla 17 que contiene las características más representativas de los afiliados en estudio.

**Tabla 17. Resumen Análisis de Conglomerados Jerárquicos 1,2 y 3.**

Conglomerado 1				Conglomerado 2				Conglomerado 3			
Edad	No.	%		Edad	No.	%		Edad	No.	%	
27 y 37	89	60,1%		27 y 37	159	63%		27 y 37	63	62,38%	
38 y 49	50	33,8%		38 y 49	80	32%		38 y 49	35	34,65%	
> = 50	8	5,4%		> = 50	8	3%		16 y 26	3	2,97%	
16 y 26	1	0,7%		16 y 26	4	2%		<b>Total general</b>	<b>101</b>	<b>100,00%</b>	
<b>Total general</b>	<b>148</b>	<b>100,00%</b>		<b>Total general</b>	<b>251</b>	<b>100%</b>					
Regional	No.	%		Regional	No.	%		Regional	No.	%	
NORTE	136	91,89%		CENTRO	150	60%		NORTE	33	32,67%	
OCCIDENTE	12	8,11%		ANTIOQUIA	65	26%		CENTRO	32	31,68%	
<b>Total general</b>	<b>148</b>	<b>100,00%</b>		OCCIDENTE	36	14%		OCCIDENTE	23	22,77%	
				<b>Total general</b>	<b>251</b>	<b>100%</b>		ANTIOQUIA	13	12,87%	
								<b>Total general</b>	<b>101</b>	<b>100,00%</b>	
Tipo de empresa	No.	%		Tipo de empresa	No.	%		Tipo de empresa	No.	%	
Alto	107	72,30%		Alto	136	54%		Bajo	70	69,31%	
Superior	41	27,70%		Superior	110	44%		Medio	25	24,75%	
<b>Total general</b>	<b>148</b>	<b>100,00%</b>		Medio	5	2%		Alto	6	5,94%	
				<b>Total general</b>	<b>251</b>	<b>100%</b>		<b>Total general</b>	<b>101</b>	<b>100,00%</b>	
IBC	No.	%		IBC	No.	%		IBC	No.	%	
535.600 Y 1.071.200	123	83,11%		535.600 Y 1.071.200	173	69%		535.600 Y 1.071.200	61	60,40%	
1.071.201 Y 2.142.400	13	8,78%		2.142.401 Y 4.284.800	29	12%		2.142.401 Y 4.284.800	16	15,84%	
535.600 Y 1.071.201	7	4,73%		1.071.201 Y 2.142.400	28	11%		>=4.284.801 Y 8.569.60	14	13,86%	
2.142.401 Y 4.284.800	3	2,03%		>=4.284.801 Y 8.569.60	21	8%		1.071.201 Y 2.142.400	10	9,90%	
>=4.284.801 Y 8.569.60	2	1,35%		<b>Total general</b>	<b>251</b>	<b>100%</b>		<b>Total general</b>	<b>101</b>	<b>100,00%</b>	
<b>Total general</b>	<b>148</b>	<b>100,00%</b>									

Las diferencias más representativas se describen a continuación:

- En el conglomerado 1 se encuentran con mayor representación los afiliados de la Regional Norte (91,89%), mientras que en conglomerado 2 los afiliados de la Regional Centro (60%) y en el conglomerado 3 los afiliados se encuentran distribuidos en las diferentes Regionales con porcentajes que oscilan (32,67% y 12,87%).
- En el conglomerado 1 un alto porcentaje de afiliados trabajan en un tipo de empresa Alto (72,30%), en el conglomerado 2 tipo de empresa Alto (54%) y

Superior (44%), a diferencia del conglomerado 3 tipo de empresa Bajo (69,31%).

- En el conglomerado 2 (20%) y 3 (29,7%) un alto porcentaje de afiliados ganan más de 2.142.401, a diferencia del primer conglomerado con un (3,38%).

## 9. CONCLUSIONES

El análisis de correspondencia múltiple permitió verificar que el 56,9% de la varianza se explica con las 2 dimensiones del modelo y su Alfa de Cronbach corresponde al 49,8%.

El método de clasificación por conglomerados jerárquicos que más se ajustó fue el método de Ward, de donde se escogió la solución con 3 conglomerados ya que se comparó el error cuadrático medio con las demás técnicas de clasificación. En este método se pudo observar que el ECM fue de 0,69940, éste presentó la mejor distribución de los datos, de acuerdo al criterio de los expertos por concentrar el mayor número de afiliados en los conglomerados y se obtuvo la siguiente distribución: un 30% el conglomerado 1, 50% el conglomerado 2 y un 20% el conglomerado 3, este presentó buenas características, ya que los conglomerados que maneja suelen ser muy compactos y persigue la minimización de la varianza intragrupal y maximiza la homogeneidad dentro de los grupos.

La técnica de análisis jerárquicos por dendograma permitió comparar los conglomerados, definir los grupos y observar cómo se relacionan las variables cualitativas dentro y entre los grupos, clasificándolos en 3 conglomerados, en el conglomerado 1 se encuentran los afiliados con edades entre 27 y 37 años (60,10%), vinculados en la Regional Norte (91,89%), tipo de empresa alto (72,30%) e IBC entre 535.600 y 1.071.200 (83,11%), en el conglomerado 2 se encuentran los afiliados entre 27 y 37 años (63%), vinculados en la Regional Centro (60%), trabajan en un tipo de empresa Alto (54%), presentan un IBC entre 535.600 y 1.071.200 (69%) y un IBC entre 2.142.401 y 4.284.000 (12%) y en el conglomerado 3 se encuentran los afiliados entre 27 y 37 años (62,38%), vinculados en la Regional Norte (32,67%), Regional Centro (31,66%), tipo de

empresa donde labora Bajo (69,31%), presentan un IBC entre 535.600 y 1.071.200 (60,40%) y un IBC entre 2.142.401 y 4.284.000 (15,84%).

Variable	SEGMENTO 1		SEGMENTO 2		SEGMENTO 3	
Edad	27 Y 37 años	60,10%	27 Y 37 años	63%	27 Y 37 años	62,38%
Regional	Norte	91,89%	Centro	60%	Norte	32,67%
					Centro	31,66%
Tipo de empresa	Alto	72,30%	Alto	54%	Bajo	69,31%
IBC	535.600 y 1.071.200	83,11%	535.600 y 1.071.200	69%	535.600 y 1.071.200	60,40%
			2.142.401 y 4.284.000	12%	2142.401 y 4.284.000	15,84%

## 10. DISCUSIÓN

Teniendo en cuenta la segmentación obtenida de la implementación de la técnica de clasificación por conglomerados jerárquicos, se han planteado las siguientes categorías, de acuerdo a las estrategias que existen actualmente en el fondo de pensiones obligatorias, las cuales permitirán diseñar el modelo de segmentación y mercadeo y de esta forma reducir el riesgo de deserción, aumentar la fidelización y satisfacción de los afiliados.

**Tabla 18. Estrategias Comerciales**

Conglomerado	Estrategia
1	Prioridad en PQRS
	Brochure electrónico
	Envío de extractos según la ley
	Recepción de solicitudes Contact Center
	Recepción de tramites Oficinas
2	Gestión proactiva de retención
	Gestión comercial de satisfacción
	Prioridad en PQRS
	Brochure informativo de ventajas
	Envío de extractos según la ley
	Recepción de solicitudes Contact Center
	Recepción de tramites Oficinas
3	Envío de extractos según la ley
	Recepción de solicitudes Contact Center
	Recepción de tramites Oficinas

La técnica tiene una gran capacidad de agrupar afiliados por su similaridad, lo que permite direccionar las estrategias con un uso eficiente en la administración de la información, de manera que se constituye en una valiosa herramienta en el proceso de retención, fidelización o recuperación de clientes; logros que se confirmarán una vez sean implementadas las estrategias.



## **11. LISTA DE REFERENCIAS**

### **BIBLIOGRAFÍA**

Cómo aplicar un clúster jerárquico en SPSS - Ruth Vilá-Baños, María-José Rubio-Hurtado, Vanesa Berlanga-Silvente, Mercedes Torrado- Fonseca.

Departamento de Capacitación, (2012), Cartilla del participante y estrategia corporativa pensión obligatoria, Cartilla de Inducción, (2-10,20-25).

(Dirección Nacional de Capacitación, 2005). La fórmula del conocimiento Ley 100 de 1993. Bogotá: Colfondos s.a.

Gerencia Inteligencia de Negocios (2012), análisis y referencia estratégica corporativa, 1(2), 35-52, Bogotá, Colfondos s.a.

Universidad de Córdoba, Análisis clúster, Prof. Dr. José Perea, Dpto. producción.

## WEB GRAFÍA

[es.wikipedia.org/wiki/Segmentación\\_de\\_mercado](http://es.wikipedia.org/wiki/Segmentación_de_mercado).

<https://www.porvenir.com.co/Personas/PensionesObligatorias/AcercaProducto/>.

[www.wordreference.com/definicion/linealidad](http://www.wordreference.com/definicion/linealidad)

[es.wikipedia.org/wiki/Condición\\_de\\_colinealidad](http://es.wikipedia.org/wiki/Condición_de_colinealidad).

Definición. DE (2013), internet: <http://definicion.de/previsional/>

Word Reference.com (2014) internet,

<http://www.wordreference.com/definicion/predecir>

<http://www.superfinanciera.gov.co/Cifras/financiera/pensiones/ahorroindividual/estadisticas/afpo2013.xls>

Definición. DE (2013), internet: <http://definicion.de/previsional/>

Wikipedia (2014), internet: <http://es.wikipedia.org/wiki/Wikipedia:Portada>

Guía para la interpretación de los dendogramas:

<http://www.uoc.edu/in3/emath/docs/Cluster.pdf>

<http://www.wordreference.com/sinonimos/aglomerar>

<http://es.wikipedia.org/wiki/Dendrograma>

<http://tisconsulting.org/es/news/market-segmentation-basic-strategies/>

<http://www.fuenterrebollo.com/Economicas/Econometria/Reducirdimension/Correspondencias/correspondencias.pdf>

<http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/GuiaSPSS/22conglj.pdf>

<http://www.webmining.cl/2011/01/proceso-de-extraccion-de-conocimiento/>